

## EC2010 - Computer Assignment #1

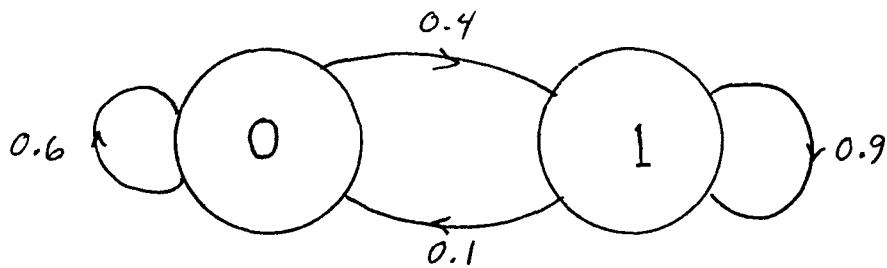
### CONDITIONAL PROBABILITY AND STATISTICAL INDEPENDENCE

#### Objective:

To study independence and conditioning in a binary data stream.

#### Procedure:

1. Simulate and plot a random sequence of 0's and 1's where each successive digit is chosen independently with  $\Pr[0]=\Pr[1]=0.5$ . Let's call this Case A. Plot a sequence for  $N=50$  successive digits. Repeat the experiment; plot another set of 50 digits and demonstrate that it is different from the first.
2. Repeat step 1 but now assume that successive digits are independent with probabilities  $\Pr[1]=0.8$  and  $\Pr[0]=0.2$ . Call this Case B. Again plot two realizations of the experiment. The model of binary data used in cases A and B is known as a ***Bernoulli process***.
3. Now let us assume that the digits are not independent; rather, the probability of a 1 or a 0 depends on whether the previous digit was a 1 or a 0. Such a sequence is known as a ***Markov chain***. We will call this Case C. The Markov chain can be represented by a state diagram as shown below:



Here the states indicate whether the previous digit was a 1 or a 0 and the numbers on the branches are conditional probabilities. For example, the probability of a 0 given that the previous digit was a 0 is 0.6. You can assume that the probability of a 0 or a 1 for the first digit in the sequence is the same as in case B. Plot a sequence of  $N=50$  digits for this Markov chain model of binary data. Repeat this; plot another realization of 50 binary digits.

(Continued on reverse side of this page.)

4. Now generate sequences of  $N$  points for each of cases A, B, and C and for the values of  $N$  specified below. Do not turn in plots these sequences. For each sequence generated, estimate the probability of a 0 or 1 in the sequence by computing the **relative frequency** of 0's and 1's. That is, for a given sequence of length  $N$  the probability of a 1 is the number of 1's appearing in the sequence divided by  $N$ . For each of the cases A, B, and C, compute the relative frequency of 1's and 0's. Summarize your results in a table for  $N=50, 100, 500, 1000$ , and  $5000$ . \_

Are the relative frequencies of 1's and 0's in any of cases A, B, and C very similar?

If so, what is the difference in the characteristics of the sequences generated by these cases? (You may want to look at plots (on your display monitor) of some of the longer sequences to better answer this question.) Discuss where each model (the Markov chain or the Bernoulli process) would be most appropriate in representing sequences of binary data occurring in the real world.

## Report

Turn in 6 plots (2 for #1, 2 for #2, 2 for #3), the table referred to in #4 and your answers to questions in #4; plus your code.